# IMAGE WATERMARKING ROBUST AGAINST NON-LINEAR VALUE-METRIC SCALING BASED ON HIGHER ORDER STATISTICS

*Fabrizio Guerrini*[1]*, Riccardo Leonardi*[1]*, Mauro Barni*[2]

[1]Department of Electronic for Automation, University of Brescia
[2] Department of Information Engineering, University of Siena
e-mail: {fabrizio.guerrini, riccardo.leonardi}@ing.unibs.it, barni@dii.unisi.it

## ABSTRACT

A new QIM-based image watermarking system for still images is proposed. The new system is expressly designed to cope with non-linear value-metric scaling attacks such as histogram stretching and gamma correction. By recognizing that any value-metric scaling attack must not change the global appearance of the image, we argue that the watermark should be inserted into high level visual features. We move a first step into this direction by proposing a system embedding the watermark into the kurtosis of selected image blocks. Though the kurtosis is not strictly invariant against non-linear gain, its value tends to remain constant whenever the image content is not altered significantly. The experiments we carried out confirm the validity of the new system, though some problems still need to be solved to make it suitable for real applications.

## 1. INTRODUCTION

The development of watermarking algorithms based on the QIM paradigm [1] that are robust against value-metric scaling is an active research field. It is well known, in fact, that weakness against value-metric scaling is one of the main drawbacks of QIM schemes with respect to classical spread spectrum algorithms [2]. In the last years, many solutions to this problem have been proposed including adaptive quantization DM [3], dirty-trellis watermarking [4], watermarking based on orthogonal-codes [5], and also Rational Dither Modulation (RDM) [6]. An approach which is somewhat similar to those described in [3, 6] was also proposed by Mihcak et al in [7]. The common denominator of all the works proposed so far is that they focus on the so-called constant gain attack, whereby the features hosting the watermark are multiplied by a constant gain factor unknown to the decoder. Despite the importance of the constant gain attack, other classes of value-metric attacks must be considered, such as, for instance, non-linear scaling and space- (time-) varying scaling; in particular, this paper focuses on the former class of attacks.

Applying a non-linear gain to pixel grey levels is a rather common operation. For instance, it may be used to correct the non-linearities of CRT display (gamma correction), to augment the image contrast (histogram stretching) or to improve the overall *readability* of the image by lightening dark areas and making bright pixels darker. Being such operations so common, it is mandatory that a robust watermarking algorithm survives them to a large extent. In order to exactly define the non-linear gain attack, let us indicate the vector with the to-be-marked grey levels by $\mathbf{x} = \{x_1, x_2 \dots x_n\}$ and the corresponding watermarked values by $\mathbf{y} = \{y_1, y_2 \dots y_n\}$. Robustness against non linear scaling requires that the detector is able to reveal the watermark presence into an attacked version of $\mathbf{y}$ obtained as:

$$\mathbf{z} = g(\mathbf{y}), \tag{1}$$

where $g(\cdot)$ is a generic non-linear function applied point-wise to all the image pixels, i.e.,

$$z_i = g(y_i), \quad i = 1, 2 \dots n. \tag{2}$$

In the following section we propose a general approach to cope with the above attack. Then, in section 3, we introduce a specific algorithm designed to work with still images in the presence of histogram stretching and gamma correction. Some experimental results demonstrating the validity of the proposed system are given in section 4. Conclusions and directions for future work are drawn in section 5.

## 2. OVERVIEW OF THE GENERAL APPROACH

In this paper we consider 1-bit watermarking, where the detector is only asked to verify the presence of the watermark. We use a QIM approach wherein the presence of the watermark is determined by verifying whether the marked features are close enough to the quantization centroids. The easiest way to design a QIM watermarking algorithm that survives the general attack described in equations (1) and (2) consists in extracting from the vector $\mathbf{x}$ a new feature $f$ that is invariant with respect to $g$ and then applying a standard (scalar) QIM algorithm to $f$. In other words we would need to find a function $f$ such that:

$$f = f(\mathbf{x}) = f(g(\mathbf{x})), \tag{3}$$

and then apply a QIM algorithm, e.g., DM or DC-DM [1], to $f$. Of course, due to the generality of the function $g$, this is an impossible task. Fortunately, some important restrictions apply to the function $g(\cdot)$, since the processed version $\mathbf{z}$ of the marked image must be perceptually equivalent to $\mathbf{y}$. For instance, the function $g(\cdot)$ could be required to preserve the order of pixel values, hence restricting the analysis to monotonic functions. Even so the class of admissible $g(\cdot)$ is too large, and the problem of finding a feature $f$ invariant to all the admissible $g(\cdot)$ appears to be very complicated.

The approach we proposed here is slightly different. By recognizing that any admissible function $g(\cdot)$ must preserve the visual appearance of the image, we argue that choosing a feature $f$ which is related to the semantic content of the host image is sufficient to ensure robustness (if not complete invariance) against the presence of $g(\cdot)$. Now the problem is to identify a good semantic feature that can be easily quantized to embed the bit $b$. Note that classical semantic features such as image segments or edges may not satisfy this requirement for the difficulties of modifying them in a predefined way.

The solution that we explore in this paper is to use normalized higher order statistics as to-be-quantized features. This choice is motivated by observing that such features are related to the shape of the histogram of the host image, and that the histogram somewhat reflects the image content. For instance, a bimodal histogram (corresponding to a high fourth order moment) is a clue that the image consists of two distinct regions having a different gray level. To be specific, according to the results of some preliminary tests we carried out on a set of 100 images, we found that the kurtosis $\beta_2$ (normalized fourth order central moment, which is invariant to affine transformations) exhibits a poor sensitivity to two important classes of operations, namely gamma correction and histogram stretching (see figure 1).

The kurtosis of a given $\mathbf{x} = \{x_1, x_2 \dots x_n\}$ is defined as:

$$\beta_2(\mathbf{x}) = \frac{\mu_4(\mathbf{x})}{\mu_2^2(\mathbf{x})} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^4}{(\sum_{i=1}^{n}(x_i - \overline{x})^2)^2}, \tag{4}$$

where $\mu_i$ is the $i$-th order central moment and $\overline{x}$ is the mean of $\mathbf{x}$.

Watermarking is achieved by quantizing the kurtosis feature. Quantization is achieved by modifying the values of $\mathbf{x}$ in such a way that the kurtosis assumes the desired value, while minimizing the introduced distortion.

## 3. WATERMARKING SYSTEM DESCRIPTION

In this section we describe in detail the watermarking algorithm that we developed by starting from the considerations given in the previous section. The watermark embedding system proposed is depicted in figure 2. First, the host image is wavelet decomposed. The second-level approximation coefficients of a 8-length Daubechies wavelet decomposition is the chosen watermark domain, both for its low-pass and spatial localization characteristics and for dimensionality reduction. Then the wavelet band is subdivided into $N$ square blocks on a regular basis; in our case, for a 512×512 pixels image, $K$ =256 square blocks of 8×8 coefficients are obtained. For every block, the coefficients kurtosis is evaluated.

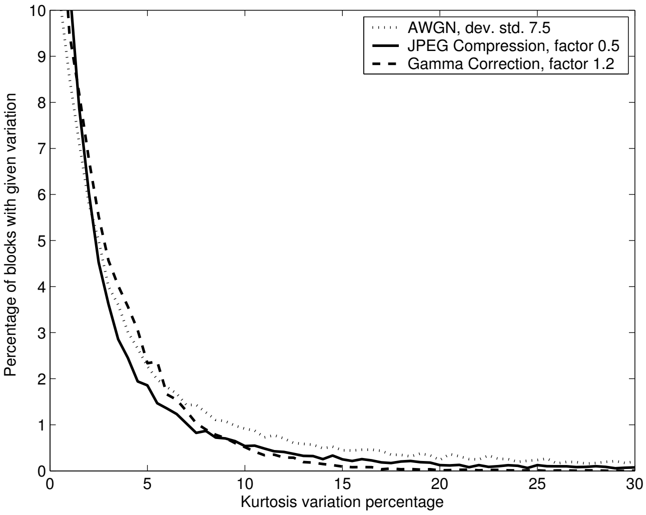
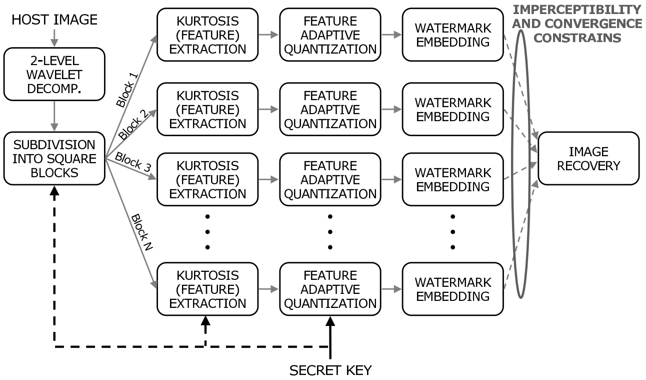


**Fig. 1**. Percentage of 8 × 8 blocks whose absolute variation is the percentage specified by the x-axis

As suggested by figure 1, we employ an adaptive quantization of the kurtosis quantization, i.e. we use 3 different quantization steps $\{\Delta_1, \Delta_2, \Delta_3\}$, where the larger steps are used for higher kurtosis values. Furthermore, we define a range of *embeddable* kurtosis so to avoid very low kurtosis values because they tend to blow successive watermark embedding and very high values for their erratic behavior (i.e. tendency to vary considerably). We derive randomly from the secret key a length-$N$ vector $\mathbf{d}$ and we shift the reconstruction values of the quantizer for the $n$-th block by $d_n$, with $-\Delta_i/4 \leq d_n \leq \Delta_i/4$. The quantization rule in the non-shifted case employed by the embedding system is depicted in figure 3. The successive step is watermark embedding in



HOST IMAGE
2-LEVEL WAVELET DECOMP.
SUBDIVISION INTO SQUARE BLOCKS
Block 1
Block 2
Block 3
Block N
KURTOSIS (FEATURE) EXTRACTION
FEATURE ADAPTIVE QUANTIZATION
WATERMARK EMBEDDING
IMPERCEPTIBILITY AND CONVERGENCE CONSTRAINS
IMAGE RECOVERY
SECRET KEY

**Fig. 2**. Watermarking embedding flow-chart the wavelet domain. Given the host coefficients vector $\mathbf{x}$ with kurtosis $k$ we wish to find a new (watermarked) coefficients vector $\mathbf{y}$ constrained to have kurtosis $h$, the quantized version of $k$. We search the solution of this under-determinated problem that minimizes the $L_2$ norm of the watermark $\mathbf{n} = \mathbf{y} - \mathbf{x}$. This non-linearly constrained minimization of a non-linear function is an operation computationally expensive and prone to divergence. Hence, we slightly relax the constrain and solve iteratively what follows:

$$\min_{\mathbf{y}} \|\mathbf{n}\| \quad \text{subject to} \quad h - \epsilon \le \beta_2(\mathbf{y}) \le h + \epsilon \qquad (5)$$

where $\epsilon$ is a small tolerance parameter. Finally, the watermarked image is recovered from the watermarked approximation coefficients, obtained substituting $\mathbf{x}$ with $\mathbf{y}$ for every block, and from the host detail subbands.

Solving equation (5) leads to a number of problems. As it is a strongly non-linear problem, it might diverge; even when it converges to a solution, it could introduce in the pixel domain visible artifacts or saturated pixel. In case of divergence, the block is left untouched. The visibility problem is addressed evaluating the distortion introduced in the image; if it is excessive, again the block is left untouched. Dashed lines in figure 2 account for the possible discarding of watermarked blocks. The saturation problem is critical because storing the watermarked image in a file clips the saturated pixel values and could significatively modify the feature value; to avoid this problem, we have so far considered artificially reduced dynamic range images.

Observe that leaving unmarked blocks inside the decoding range affects system performance; in fact they have a 0.5 probability of decoding error because of the uniform pdf of $\mathbf{d}$. In an attempt to reduce the visibility and/or convergence problem, the embedder tries to reach some nearer feature value, obviously under the same visibility constraints, up to the border of the correct decoding region.

Another problem is given by the occurrence of unpredicted decoding of unmarked blocks whose feature value has entered the decoding range after some kind of image processing. To reduce these events, the embedder makes an effort to move blocks with out-of-range kurtosis value well away from the decoding range.

The detection process is similar to classical QIM; the decoder evaluates the kurtosis feature for every block in the same fashion of the embedder. If the feature $h'$ is inside the decoding range, it performs the adaptive quantization by using the key-derived $\mathbf{d}$ vector and takes a binary decision whether the $n$-th block is marked as stated in equation 6.

$$h' \quad \text{is marked if} \quad |Q(h', \Delta_i, d_n - h'| \le \Delta_i/4 \qquad (6)$$

where $Q(h', \Delta_i, d_n)$ represents the quantization with step $\Delta_i$ and shift $d_n$ and $\Delta_i$ is derived from the decoding region in which $h'$ lies. To decide whether an image is watermarked or not, the decoder compares the number of blocks correctly decoded with a threshold $T$, i.e. an assigned percentage of the decoded blocks that realizes a suitable tradeoff between false alarm probability and miss probability.

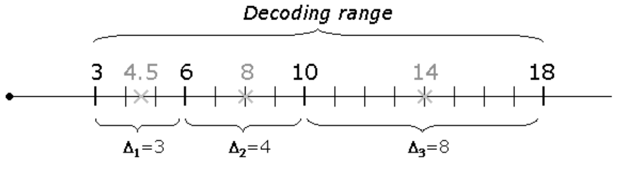


**Fig. 3**. Non-shifted quantization rule. Subtract $d_n$ to obtain the quantization rule for the $n$-th block

## 4. EXPERIMENTAL RESULTS

In order to evaluate robustness performances, we derived the Receiver Operating Characteristics under various kinds of image processing: addition of white Gaussian noise, gamma correction, histogram stretching, JPEG compression and constant gain attack. To estimate the expected very low miss and false alarm probability we operated this way: we watermarked 100 512×512 pixels images and for every processed watermarked image we evaluated a mean block error probability $p_e$ from the resulting 25600 potential decodings (including untouched blocks); given the threshold $T$, the miss probability $p_m(T)$ is then derived as the probability of having at least $\lceil T \cdot N \rceil$ correct block decodings (where $N$ is the minimum overall block decodings performed on the images):

$$p_m(T) = \sum_{i=\lceil T \cdot N \rceil}^{N} \binom{N}{i} (1 - p_e)^i \cdot p_e^{N-i} \qquad (7)$$

As previously stated, unmarked blocks have 0.5 block error probability; hence the false alarm probability $p_{fa}(T)$ is derived by substituting $p_e = 0.5$ in equation (7), thus obtaining:

$$p_{fa}(T) = \sum_{i=\lceil T \cdot N \rceil}^{N} \binom{N}{i} 2^{-N} \qquad (8)$$

Predicted ROCs are depicted in figure 4. As expected from figure 1, the system is more sensitive to AWGN than to gamma correction, but it retains good performances for all the types of attacks considered. Constant gain attacks, as expected, result in rare block errors (due to untouched blocks).

## 5. CONCLUSIONS AND FUTURE WORK

We proposed a new QIM-based image watermarking algorithm that relies on kurtosis of wavelet approximation coefficients to achieve robustness against non-linear scaling attacks. Adaptive quantization is employed to match feature behavior. Experiments carried out so far show good performance in presence of histogram stretching and gamma correction, as well as with other common attacks such as AWGN and JPEG compression.

Open challenges include security and saturation effects. Security could be assured by inserting key-derived parameters into the block subdivision and feature evaluation steps (black dashed lines in Figure 2). In particular, block shape and position could be random (as in the Bubble Random Sampling approach [8]) and coefficients might be randomly weighted (as in [7]) before evaluating the kurtosis. Saturation effects might be mitigated by a post-processing step similar to the one used for the convergence problem.

## 6. REFERENCES

[1] B. Chen and G. Wornell, “Quantization index modulation: a class of provably good methods for digital watermarking and information embedding,” *IEEE Trans. on Information Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[2] F. Bartolini, M. Barni and A. Piva, “Performance analysis of ST-DM watermarking in presence of non-additive attacks,” *IEEE Trans. on Signal Processing*, vol. 52, no. 10, pp. 2965–2974, October 2004.

[3] J. Oostven, T. Kalker and M. Staring, “Adaptive quantization watermarking,” in *Security, Steganography, and Watermarking of Multimedia Contents VI, Proc. SPIE Vol. 5306*, P. W. Wong and E. J. Delp, Eds., pp. 296–303, San Jose, CA, USA, January 2004.

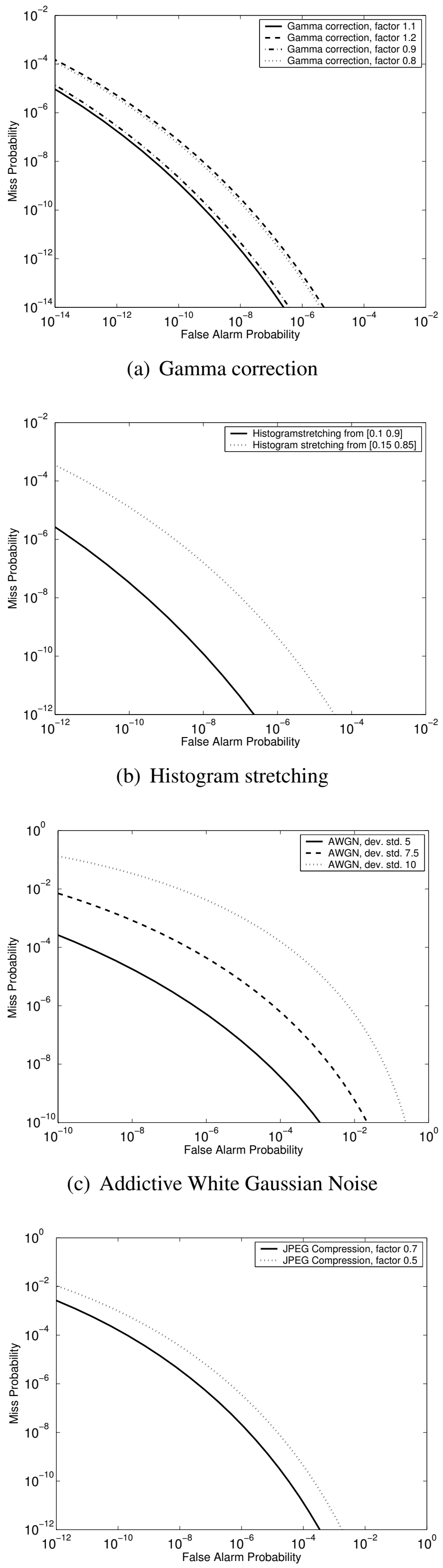
[4] M. L. Miller, G. J. Doerr and I. J. Cox, “Applying informed coding and embedding to design a robust, high capacity, watermark,” *IEEE Trans. on Image Processing*, vol. 13, no. 6, pp. 792–807, June 2004.

[5] A. Abrardo and M. Barni, “Informed watermarking by means of orthogonal and quasi-orthogonal dirty paper coding,” *IEEE Trans. on Signal Processing*, vol. 53, no. 2, pp. 824–833, February 2005.

[6] F. Perez-Gonzalez, C. Mosquera, A. Abrardo and M. Barni, “Rational dither modulation: a high-rate data-hiding method invariant to gain attacks,” *IEEE Trans. on Signal Processing*, vol. 53, no. 10, Part II, pp. 3960–3975, October 2005.

[7] K. Mihcak, R. Venkatesen and T. Liu, “Watermarking via optimization algorithms for quantizing randomized semi-global image features,” *to appear in Special Issue about Multimedia Security for ACM Multimedia Systems Journal*, 2005.

[8] F. Guerrini, R. Leonardi and P. Migliorati, “Image retrieval with random bubbles,” in *Proc. EUSIPCO*, pp. 1035–1038, Wien, Austria, September 2004.



(a) Gamma correction

(b) Histogram stretching

(c) Addictive White Gaussian Noise

(d) JPEG Compression

**Fig. 4**. Receiver Operating Characteristics (ROCs) under different image processing